DATE: 4/01/97

CONTROLLING OFFICE FOR THIS DOCUMENT IS:
DIRECTOR, Army High Performance Computing
Research Center (AHPCRC)
Army Research Laboratory
Aberdeen, MD 21005

POC: Director (Tayn E. Tezduyar)

DISTRIBUTION STATEMENT A: Public release

Computational Study on Hydrogen Bonding and Stacking Interactions Between Nucleic Acid Bases

Sum 95

by Jerzy Leszczynski (AHPCRC-JSU)

Theoretical calculations play an ever-increasing role in studies of molecular structures, properties, and chemical reactions. Leszczynskis main research interests are concerned with the study, by use of nonempirical ab initio orbital calculations, of the structures and stabilities of various chemical species and the details of their intermolecular interactions. He uses high-level ab initio methods both to rationalize existing chemistry and to predict new species characterized by exciting chemical properties.

Computational chemistry is a powerful, though computationally demanding tool, for the investigation of chemical systems. When such calculations are carried out at the appropriate level (ab initio calculation with large basis sets and with inclusions of electron correlation contributions), the quality of the predicted molecular bond distances and angles, dipole moments, IR harmonic frequencies, and activities is comparable to that of those obtained by accurate experimental techniques. Due to a vast increase in the numbers of computing platforms and efficient high performance computational algorithms in recent years,such calculations have become possible even for large molecules of biological importance. Since 1990, theoretical ab initio studies on nucleic acid bases, their modified analog and hydrogen-bonded and stacked DNA base pairs have become one of the major research areas for Professor Leszczynskis group. Successful continuation of these investigations has to a large extent been possible due to consecutive support from the AHPCRC and access to its computer facilities. An important component of these facilities is the Molecular Modeling Lab at JSU, which was initially funded by the AHPCRC and later upgraded by DoD and NSF awards. Also the Lab and the computational group constantly benefit from interactions with and support of the staff of the Waterways Experiment Station, which is located 30 miles west of Jackson.

Nucleic acid bases are the basis components of the DNA and RNA structures. For many years, isolated DNA bases were considered to be planar. Due to a lack of experimental data for the DNA base amino groups, such assumptions were derived from approximated empirical force-field calculations, which penalize non-planar geometries of the -NH2 groups. The Leszczynski Groups first high-level comprehensive study on cytosine, isocytosine, 1-methylcytosine, adenine, and guanine proved that, in contrast to empirical force fields, ab initio calculations predict non-planar geometries for the amino groups of the investigated bases. Such

non-planar structures and deformability of the DNA amino groups allow explanation of many conformational properties and base-base interactions in biological systems.

Hydrogen bonding of the DNA base is important for the stability of the DNA double helix and it plays critical role in providing specificity for information transfer. For the most part, guanine is paired with cytosine via three hydrogen bonds, and adenine is complexed with thymine through two hydrogen bonds. In addition to these standard Watson-Crick base pairs, many other base pairs contribute to the conformational variability of DNA. Stacking and hydrogen-bonding interactions of DNA bases represent an important source of conformational variability of DNA.

Hydrogen bonding in DNA bases was investigated using reliable nonempirical ab initio computational methods. Gradient optimization was carried out on 30 DNA base pairs using the Hartree-Fock (HF) approximation and the 6-31G** basis set. The optimizations were performed within the Cs symmetry. However, the harmonic vibrational analysis indicates that 13 of the studied base pairs are intrinsically somewhat nonplanar. The interaction energies of the base pairs were then evaluated at the optimized planar geometries with inclusion of the electron correlation energy at the MP2 level. The stabilization energies of the studied base pairs range from 24 kcal/mol to 9 kcal/mol, and the calculated gas-phase interaction enthalpies agree well (within 2 kcal/mol) with the available experimental values. The binding energies and molecular structures of the base pairs are not determined solely by the hydrogen bonds, but are also strongly influenced by the polarity of the monomers and by a wide variety of secondary long-range electrostatic interactions involving the hydrogen atoms bonded to ring carbon atoms. The stabilization of the base pairs is dominated by the Hartree-Fock interaction energy. This result confirms that the stability of the base pairs originates in the electrostatic interactions. For weakly bonded base pairs, the correlation interaction energy amounts to as much as 30-40% of the stabilization. For some other base pairs, however, a repulsive correlation interaction energy was found. This fact is explained as a result of a reduction of the electrostatic attraction upon inclusion of the electron correlation. The empirical London dispersion energy does not correctly reproduce the correlation interaction energy. For the sake of comparison, results of a first gradient optimization for a DNA base pair at a correlated level (CC base pair, MP2/6-31G** level) are reported. In addition, the ability of the economical Density Functional Theory (DFT) method to reproduce the ab initio data was investigated. The DFT method with its presently available functionals is not suitable for a consistent study of the whole range of the DNA base interactions. However, it gives good estimates of interaction energies at the reference HF/6-31G** geometries.

The Leszczynski Group has also investigated the energetic provisions for Lowdins DNA mutational mechanism (Lowdin, P. O.

Rev. Mod. Phys. 35, 724 (1963)) of the formation of substitution DNA mutations in the guanine-cytosine Watson-Crick base pair. The structures studied involve the canonical base pair (GC1), rare base pair tautomers that are formed from GC1 by the antiparallel simultaneous transfer of two protons in hydrogen bonds, and ion-pair G-C+ structures that are formed by the transfer of a single proton from guanine to cytosine. The geometries of these complexes were optimized by ab initio Hartree-Fock calculations using the 6-31G* basis set. At the same level, harmonic vibrational frequencies were determined. Nonplanar geometries featuring considerable propeller-twist angles and a pyramidal guanine amino group were found for base pairs involving guanine anion and 6-hydroxyguanine. The relative stabilities and dissociation energies of the base pairs were determined at the higher MP2/6-31G**//HF/6-31G* level of theory. These methods were also used to locate transition states on the potential energy surface of this system. Starting from the geometries of two different transition states lying close to the ion-pair minimum, the intrinsic reaction coordinate for the single-proton transfer from the canonical to the 6-hydroxyguanine.4-iminocytosine tautomer (GC2) was evaluated. It was concluded that, in contrast to the adenine.thymine base pair (for which Lowdins mutational mechanism is not supported by the present theoretical data), the GC1>GC2 tautomeric transition is likely to occur at a rate of one per 106 109 guanine-cytosine base pairs. This frequency is significant from the point of view of the fidelity of DNA replication.

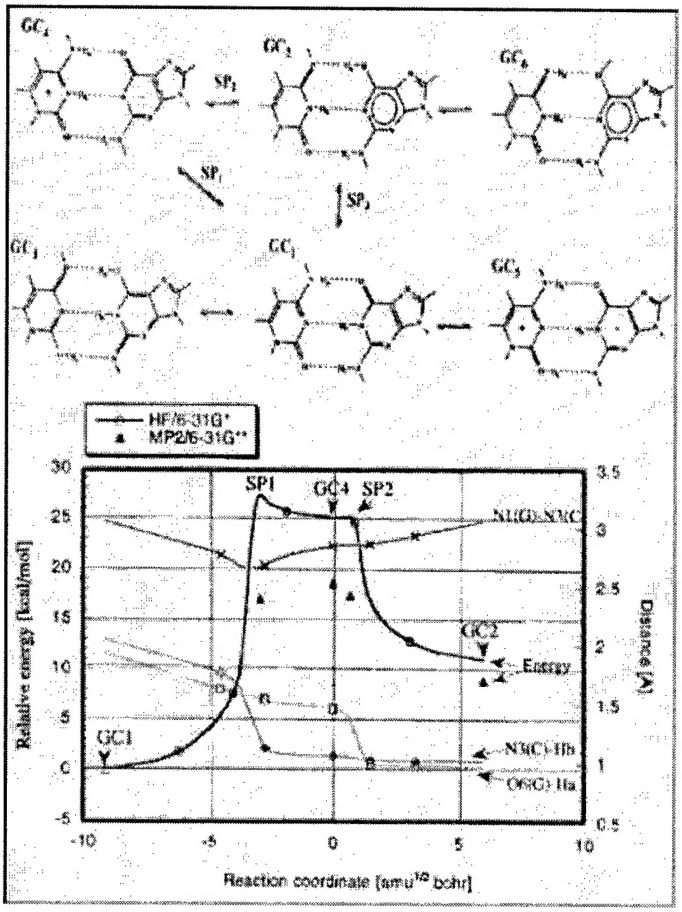

Figure 1. Top: Structure and numbering of the canonical guanine cytosine base pair. Bottom: Plot of the energy profile, hydrogen, and intermolecular bond lengths along the HF/6-31G* minimum energy path (MEP,IRC) connecting GC,GC4 and GC2 structures.

Recently, new pi and k bases were proposed as an extension of the genetic alphabet from four to six letters (Piccinlli et al, Nature 343.33 (1990)). The structure of the nonclassical k base pair (7-methyl-oxoformycin B ... 2,4-diaminopyrimidine) was studied at the ab initio Hartree-Fock and MP2 levels using the 6-31G* and 6-31G** basis sets. The k base pair is bound by three parallel hydrogen bonds with the donor-acceptor-donor recognition pattern. By the HF/6-31G* method with full geometry optimization, the minimum-energy structure of this complex was calculated to have a 12-degree propeller twist. The linearity of hydrogen bonds is preserved in the twisted structure by virtue of the pyramidal arrangement of the k-base amino groups. The rings of both the pi and k molecules remain nearly planar. This nonplanar pi-k base pair structure is only 0.1 kcal/mol more stable than the planar (Cs) conformation. The HF/ 6-31G* level gas-phase interaction energy of k (-13.5 kcal/mol) was calculated to be nearly the same as the interaction energy obtained previously for the adenine-thymine base pair (13.4 kcal/mol) at the same computational

level. The inclusion of p-polarization functions on hydrogens, electron correlation effects (MP2/6-31G** level), and the correction for the basis set superposition error (BSSE) increase this energy to 14.0 kcal/mol.

A real challenge to computational chemists concerns the origin of stacking interactions in nucleic acid bases. Recently, ab initio MP2/6-31G*(0.25) interaction energies were calculated for almost 240 geometries of 10 stacked DNA base pairs: A...A, C...C, G...G, U...U, A...C, G...A, A...U, G...C, C...U, and G...U; in some cases uracil was replaced by thymine. The most stable stacked pair is the G...G dimer (-11.3 kcal/mol), and the least stable is the uracil dimer (-6.5 kcal/mol). The stacked pairs were also analyzed using empirical potential calculations. The corresponding H-bonded pairs were investigated at the MP2/6-31G*(0.25)//HF/6-31G** level. Their interaction energies range from -25.8 kcal/mol (G...C) to -10.6 kcal/mol (T...T). The stability of stacked pairs originates in the electron correlation, while stability of the H-bonded pairs is dominated by the HF energy. The mutual orientation of the stacked bases is, however, primarily determined by the HF contribution to the interaction energy. The orientational dependence of stacking energy is dominated by changes of twist, while displacements have a smaller influence on the stacking energy. Ab initio base-stacking energies are reproduced well by empirical potential calculations, except for the fact that the Lennard-Jones van der Waals potential does not satisfactorily reproduce the short-range repulsion for some structures. No other qualitative differences between the ab initio and empirical potential data were found. This demonstrates that some contributions previously postulated to significantly influence base stacking (induction interactions, pi-pi interactions) are negligible. Base stacking was also investigated in 6 B-DNA and 2 Z-DNA base-pair steps; their geometries were taken from oligonucleotide crystal data. The many-body correction was estimated at the HF/MINI-1 level. The HF and correlation contributions to the base-stacking energy of DNA base pair steps show rather large sequence-dependent variability, as do the intrastrand and interstrand contributions. The sequence-dependent variations of the total base-pair stacking energy are smaller, ranging from 9.9 to 14.7 kcal/mol. The range of calculated many-body corrections to the stacking energy is 2 kcal/mol. The ab initio calculations exclude the consideration that the unusual conformational properties of the Cpa(TpG) steps might be associated with attractive induction interactions of the exocyclic groups of DNA bases and the aromatic rings of bases.